

<https://helda.helsinki.fi>

Survey of Uralic Universal Dependencies development

Rueter, Jack

The Association for Computational Linguistics
2019

Rueter , J & Partanen , N 2019 , Survey of Uralic Universal Dependencies development . in
Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019) : Proceedings . , 78 ,
The Association for Computational Linguistics , Stroudsburg , Workshop on Universal
Dependencies , Paris , France , 29/08/2019 .

<http://hdl.handle.net/10138/306225>

cc_by
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Survey of Uralic Universal Dependencies development

Niko Partanen

University of Helsinki
niko.partanen@helsinki.fi

Jack Rueter

University of Helsinki
jack.rueter@helsinki.fi

Abstract

This paper attempts to evaluate some of the systematic differences in Uralic Universal Dependencies treebanks from a perspective that would help to introduce reasonable improvements in treebank annotation consistency within this language family. The study finds that the coverage of Uralic languages in the project is already relatively high, and the majority of typically Uralic features are already present and can be discussed on the basis of existing treebanks. Some of the idiosyncrasies found in individual treebanks stem from language-internal grammar traditions, and could be a target for harmonization in later phases.

1 Introduction

The Uralic languages constitute one of the major language families in the world. There are approximately 38 languages represented by seven major branches in the family tree. Only Finnish, Estonian and Hungarian are majority languages in their states, and other Uralic languages are minority languages, often endangered, in their respective regions, including Northern Scandinavia and Russia.

Uralic languages are agglutinative, morphologically rich and typically have large case systems. The majority of Uralic languages share many features, including, the expression of negation with a verb of negation, preference of postpositions and complex object marking. They usually also have a complex system of non-finite forms. Constituent order is relatively flexible. However, since the different branches of the Uralic family are rather far removed from one another, there are also numerous independent reflexes of historically shared features on the individual language level.

Recent versions of Universal Dependencies treebanks (Nivre et al., 2019) include 11 treebanks in seven Uralic languages. All in all, they represent five of the aforementioned major branches in the family. In release 2.4 the Uralic languages include:

- Erzya, 1 treebank (Rueter and Tyers, 2018)
- Estonian, 2 treebanks (Muischnek et al., 2014), (Muischnek et al., 2016)
- Finnish, 3 treebanks (Haverinen et al., 2014), (Pyysalo et al., 2015)
- Hungarian, 1 treebank (Vincze et al., 2017)
- Karelian, 1 treebank (Pirinen, 2019)
- Komi-Zyrian, 2 treebanks (Partanen et al., 2018)
- North Sámi, 1 treebank (Sheyanova and Tyers, 2017)

In addition to the languages listed above, there are at least plans for a Northern Mansi treebank outlined in the literature (Horváth et al., 2017, 63), and work for Livvi Karelian treebank is undergoing. At the moment, the branches of the Uralic language family that do not have a single treebank are Samoyed and Mari, while the Ugric branch is only represented by Hungarian, missing the Ob-Ugric languages. From this point of view, the Northern Mansi treebank mentioned would be a most welcome addition. This would also bring improvement to the geographical limitations in the current selection of languages, where all now available treebanks represent Uralic languages spoken in Europe, with no language spoken primarily in Siberia.

Finite-State Transducers, especially in the Giellatekno infrastructure (Moshagen et al., 2014), have traditionally played an important role in open-source language technology for Uralic languages. Recent work in harmonizing NLP solutions (Hämäläinen, 2019) and lexical resources (Hämäläinen and Rueter, 2018) in Uralic languages may also prove beneficial for these efforts. Many of the Uralic treebanks have been created by automatic conversion of annotation schemes and tagsets from these analysers. Since several language documentation projects have also started to integrate these tools into their workflows (Gerstenberger et al., 2016), it would be expected that the pipelines used thus far could be reused when creating new treebanks for languages that have comparable resources.

Variation in annotation schemes has already been identified as one problem between analysers of different Uralic languages, and an earlier study by Tyers and Pirinen (2016) which examined this in detail. To continue research along these lines, we reproduce their Table 2 with minor corrections, and compare how the same examples align with the solution currently found in UD treebanks. This is one method to measure whether the common guidelines proposed in Tyers and Pirinen (2016) have taken root.

2 Selected Uralic features

In this section we go through selected features in the Universal Dependencies annotation scheme that can be commented specifically from a Uralic point of view.

2.1 Feature Evident

Komi-Zyrian has morphologically marked evidential forms of the verbs, and the Komi treebanks also use this feature with the value ‘Evident=Nfh’ to mark second past tense forms as non-firsthand information, although evidentiality is not there as an obligatory category as in some other languages, used primarily in unwitnessed narrative or to express non-voluntary action (Leinonen, 2000). Various evidentiality related phenomena occur in the morphology of other Permic languages, Mari and Ob-Ugric, and the Samoyedic languages, which, as mentioned, are still largely missing from the UD project. It can be stated that evidentiality is one feature for which Uralic family still has much to add in the project, although in the currently included languages it does not play a central role. The Erzya treebank, it will be noted, uses the ‘Evident=Nfh’ feature for some particles connected as advmod dependencies. Erzya treebank usage may open the discussion of introducing this feature in relation to other free word forms in the UD project. The Estonian treebank UD_Estonian-EDT, for example, does not have the feature Evident, although the concept is implicitly present in ‘Mood=Qot’.

2.2 The Feature Gender

The Uralic languages do not have grammatical gender, as it were, permeating the pronoun, noun and verb systems. They do, however, have peripheral derivational elements, which are not regularly addressed, e.g. Finnish *-tar/-tär*: *ruhtinas* ‘duke’ vs *ruhtinatar* ‘duchess’, *näyttelijä* ‘actor’ vs *näyttelijätär* ‘actress’, etc. Erzya also a peripheral derivational element: *-нuзэ/-нuзэ -hize/-nize*, which, traditionally, was added to the husband’s name for indicating ‘wife of’: *Иван Ivan* vs. *Иваннызэ Ivannize*, *Гава Gava* vs. *Гаванузэ Gavanize*. A similar construction predominantly occurs in the Hungarian language *né* ‘wife of’: *István* vs. *Istvánné*, but it is not found in the Hungarian treebank. Under normal circumstances, there should be no reason to mark gender feature in Uralic treebanks. One exception could arise in situations where the treebank contains code-switch speech to Russian, under which circumstances gender marking may be present (Janurik, 2015).

2.3 The Features Animacy and Definite

Animacy is not a grammatical category in Uralic languages, but it does influence the object marking in some languages within the family, e.g. in Komi, animacy has been connected to the marking of definiteness and focus as briefly described by Fediunova (2000, 69). At present, neither animacy nor definiteness have been marked in Komi-Zyrian treebanks, but definiteness can, in principle, be deduced from possessive suffixes used to this end. In the Erzya treebank, definiteness is marked as an incremental feature of the NP head morphology – similar to Scandinavian languages – yet it is distinct from the use of posses-

sive marking. Hungarian also uses this feature, but it is collocated with the definite articles found in the language.

2.4 The Feature Aspect

At the moment, some Erzya verbs are marked with ‘Aspect=Inch’, and some Hungarian verbs are marked with ‘Aspect=Iter’. Neither language has gone beyond expressing features for specific derivation morphology. In the Northern Saami treebank, however, ‘Aspect’ is used as a means for encoding participles, i.e. the perfect participle is coded with ‘Aspect=Perf’, whereas the present participle is coded with ‘Tense=Present’. Neither Finnish, Estonian nor Komi use the Aspect feature at this point.

2.5 The Feature Number

Among the current Uralic treebanks, Northern Saami is the only one that has a dual number. The numbers used throughout are singular and plural. On the subject of number, however, there are several types to keep track of: simple *Number* is used with nominals to indicate the number of entities, *Number[Psor]* the possessor number, of course, tells us of the possessor flagged by possessive suffixes. When we arrive on the verb scene, Erzya introduces counting subject entities flagged on the finite verb with *Number[Subj]*, and object entities as well *Number[Obj]*. Hungarian introduces counting of possessor/possessee with *Number[Psee]* (see also (Vincze et al., 2017)). This is useful in Hungarian and could be feasible in any up-coming of treebank for Moksha, as well, e.g. Hungarian *kutya* ‘dog’ vs. *kutyáé* ‘the one belonging to the dog’, Moksha *пине пи́не* ‘dog’ vs. *пиненне пи́ненне* ‘one belonging to a dog’.

The ‘Number’ strategy sets a precedence for analogical regular inflectional features in Erzya and Komi-Zyrian. Where Erzya uses some of the oblique cases at both the NP level and the VP level, Komi has an operating dichotomy that distinguishes the two levels. If, for instance, the inessive Erzya *вурьсэ ви́р-се* ‘in the forest’ (derived from *вурь ви́р* ‘forest’) is taken as a premodifier in a noun phrase, Erzya morphology allows for constructions where NP head morphology is directly concatenated onto the premodifier, which might result in a form *вурьсэтнесэ ви́р-се-т-не-се* ‘in the ones that are in the forest’ (a matter of ellipsis or ‘secondary declension’ as it is also referred to in the literature. Komi-Zyrian can derive a premodifier with the same semantics of its Erzya counterpart in *вöрса вärsa* ‘in the forest’ from *вöр вär* ‘forest’, which can in turn, as a NP head, take on either copula plural morphology (*вöрсаöсь вär-са-эс* ‘[are] ones in the forest’) or noun plural morphology (*вöрсаяс вär-са-яс* ‘the ones in the forest’). Although this regular morphology for Komi premodifiers is not addressed as case morphology in the largest of Komi grammars, (Fediunova, 2000), it merits contemplation in any extensive and parallel treatment of the language family. The fact that a second plural form can be present introduces further problems.

The Komi NP premodifier derivation strategies allow for plural stems. Hence, forms, such as *вöрьяссаяс вär-jas-са-яс* ‘the ones that are in the forests’ and *вöрьяссаöсь вär-jas-са-эс* ‘[are] the ones that are in the forests’ may require regular counting and therefore a new ‘Number’: one to express the number of the NP head and the other indicate the number of the NP premodifier. In Komi, the locative *-са -sa* has further siblings in a temporal *-ся -ça*, a privative *-мöм -täm* and a propriative *-а -a*, which should not be confused with inessive *-ын -in*, caritive *-мöг -täg*, comitative *-кöд -käd* or instrumental *-өн -än* cases.

The situation seems to be similar to what is found in the Turkic languages, and the solution proposed in Çöltekin (2016) to split these words in Turkish into multiple tokens, unless they are lexicalized, would also be possible with the Uralic languages. The fact that Uralic languages do not have separate morphology analogic to the *ki* element in Turkic, however, would seem to speak against following such a lead.

The system of number marking outlined above seems to be a good starting point for all Uralic languages. It also sets the scene for a new discussion, which might draw from other morphologically rich language families and their practices of grammar description.

2.6 The Feature Case

At the moment all Uralic treebanks use traditional terms from their own grammars. Some of the terms, for example ‘superessive’, ‘sublative’ and ‘additive’ are only used in individual languages. On the one hand, this is understandable, but it begs the question as to how useful these terms actually are. Should

these names indicate functions for a given language or should they be generalized. At least, in regard to spatial cases, the cross-lingual comparability is now rather weak. Some of the cases such as ‘terminative’, however, are already present in numerous treebanks and refer to a very similar concept. Similarly, cases like ‘approximative’, although currently present only in the Komi-Zyrian treebank, will eventually be wider present once languages with this case, such as Olonets-Karelian (Livvi), Ludic or Veps, are included.

Are the names of cases with multiple divergent functions relevant? The case names in Uralic languages are often very language-specific and are not transparent. Finnish derivation practice includes a mere mnemonic letter string to indicate a derivation, e.g. ‘Derivation=Sti’, which is basically a morphological representation for deadjectival adverb derivation. Without language-specific documentation neither case names nor derivation letter strings are meaningful. How to construct documentation in the manner that allows cross-linguistic comparison is a forthcoming challenge for treebank developers. Discussions should also include analogical solutions used in other language families, such as the English ‘in’ preposition equivalent of Finnish ‘inessive’ (Germanic prepositions are not generally given names for mapping them to equivalent Finnish case functions).

3 Selected syntactic questions

In this section we discuss some of the observations that can be made about the use of specific dependency relations, and questions that arise from morpho-syntactic particularities of Erzya and Komi.

3.1 The Dependency iobj

At the moment only the Hungarian treebank uses the dependency relation ‘iobj’. In other Uralic treebanks, however, the relation ‘obl’ is used, which is illustrative of the examples shown in UD documentation for NPs with prepositional construction closely related to ‘iobj’.¹ These present languages do not have dative alternation, but any future work with Mansi or Khanty may introduce this variation. It would be reasonable to assume that any updating of the version 2.2 Hungarian treebank would involve the introduction of the ‘obl’ relation where the ‘iobj’ is now used, perhaps with a special relation subtype of ‘obl’.

In order to compare the question better and to illustrate the situation, we translated the English and French example sentence from UD documentation into different Uralic languages.

- English: give the children the toys & give the toys to the children
- French: donner les jouets aux enfants
- Estonian: andma mänguasju lastele
- Hungarian: a játékokat a gyerekeknek adja
- Finnish: antaa lapsille leluja
- Erzya: максомс налкшкетнень эйкакштнэнь
- Komi: сетны ворсанторъяс челядьлы

All example sentences above with the exception of the first English sentence are or (in the case of Hungarian) can be coded with the ‘obl’ relation due to the explicit morphological encoding of the NP head, which distinguishes, among others, the Hungarian dative case.

3.2 Copulas

The non-past identity clause involves copula morphology in all but the Komi language. While North Sámi, Estonian, Hungarian, Finnish and Karelian all have free copulas, Erzya provides examples of dependent morphology, which have elicited segmentation (locus + copula), on the one hand, and a discussion of word order issues, on the other.

The North Sámi in figure 1a is representative of word ordering typifying Finnish, Estonian and Karelian, alike – CS COPULA CC. Komi, however, does not use a COPULA, instead, it applies juxtaposition to achieve the same (see 1b).

¹<https://universaldependencies.org/u/dep/all.html#a1-u-dep/obl>

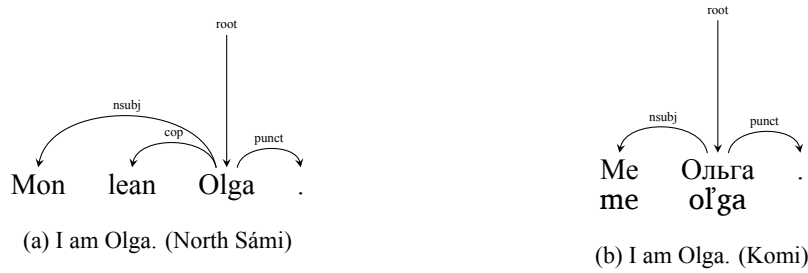


Figure 1: Example with and without copula

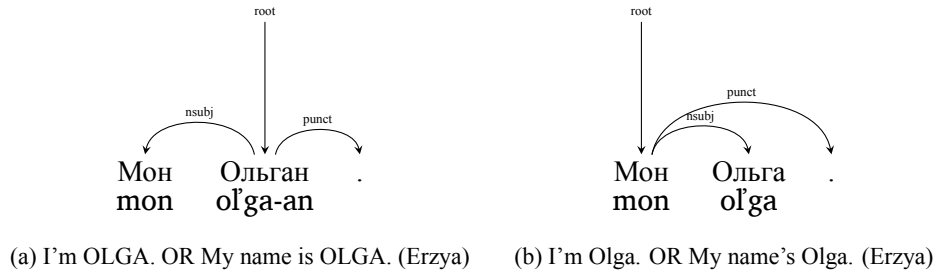


Figure 2: Distinguishing Erzya Subject

In figures (2a) and (2b), distinctive functions in Erzya morpho-syntax are presented where first and second person personal pronouns can visibly serve as both copula subjects and copula complements. The dependent copula morphology has been split off of the root in the analysis (analogic to what has been applied in the UD_Turkish-IMST treebank), but when the analysis is non-past third person singular, i.e. ZERO, no extra token is introduced (cf. Tyers and Pirinen (2016) and Vincze et al. (2017)). The logic of the split solution in Erzya can be questioned. This question is underlined by the fact that Komi-Zyrian has copula complement plural marking *-öчъ -əç*, which is used for marking attribution, location and even possession non-verbal predications non-past in much the same way as the more elaborate Erzya morphology – Komi does not split this affix off from the stem. If the copula morphology is segmented, the *-an* affix in figure (2a) is better illustrated by figure (3).

The distinction between (3) and (2b) lies in which argument the attributed agreement marking is. In (2b) it is the name that commands subject correlation, and prosodic stress falls on the personal pronoun root. In (3), however, the constant is actually the personal pronoun, and prosodic stress falls on the proper name root. These are matters, of course, for future work at discourse levels.

Unlike Turkic languages, the Erzya language has no unquestionable, distinct morphological element representing the copula in dependent marking other than what actually expresses tense, person and number. Although comparative linguistics does postulate the merging of a form of copula into the copula complement. For this reason, we are presented with a choice of following the Turkic lead, i.e. separating copula morphology from nouns (in numerous cases), adjectives and numerals for the soul purpose of reusing a ready solution. The non-past, third-person singular form of the copula complement, however, takes zero marking, which would point to non-symmetric representation of the copula construction. This, in turn, indicates a further need for a more elegant UD resolution of the issue.

4 Deverbal words and features

In Table 1 we utilize, correct and expand upon working knowledge in Uralic deverbal word constructions as represented in or analogous to UD work (cf. Tyers and Pirinen (2016)). Komi-Zyrian uses two finite clauses in sentence (i), where other languages all have a non-finite solution to the problem.

In sentence (ii), it appears that Hungarian and Estonian treat their deverbal nouns as nouns, while the other languages all encode their analogic forms with a spectrum of non-finite interpretations Conv, Ger, Inf and Sup. North Sámi, Erzya and Finnish mark person on the converb, which might not be mandatory.

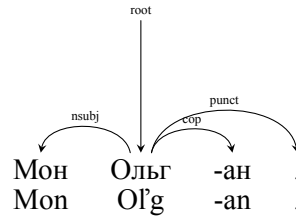


Figure 3: I'm OLGA. (Erzya)

Language	Sentence	UD
(i)	'I saw the man running.'	
North Sámi	Oidnen dievddu viehkame	Case=Ess VerbForm=Ger
Erzya	Неия цѣрантъ чїи́немадо .	Case=Abl Definite=Ind Number=Plur,Sing Valency=1 VerbForm=Vnoun
Finnish	Näin miehen juoksemassa .	Case=Ine InfForm=3 VerbForm=Inf Voice=Act
Estonian	Nägin meest juoksmas .	Case=Ine VerbForm=Sup Voice=Act
Hungarian	Láttam az embert futni	VerbForm=Inf Voice=Act
Komi-Zyrian	Аддзи, мортыс котӱрӱ .	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin
(ii)	'While running I saw the man'	
North Sámi	Oidnen dievddu viegadettiinan .	Number[psor]=Sing Person[psor]=1 VerbForm=Ger
Erzya	Чїи́немадо́ неия цѣрантъ.	Case=El Number[psor]=Sing Person[psor]=1 VerbForm=Conv
Finnish	Näin miehen juostessani .	Case=Ine InfForm=2 Number[psor]=Sing Person[psor]=1 VerbForm=Inf Voice=Act
Estonian	Jooksmise ajal nägin ma meest.	Case=Gen Number=Sing
Hungarian	Futás közben láttam az embert.	'NOUN'
Komi-Zyrian	Котралірун аддзи мортӱс.	Case=Ins Derivation=Ig Number=Sing VerbForm=Conv
(iii)	'I see the running man.'	
North Sámi	Oainnán viehkki dievddu.	Tense=Pres VerbForm=Part
Erzya	Неян чїи́ниия цѣрантъ.	Case=Nom Definite=Ind Number=Sing Tense=Pres VerbForm=Part
Finnish	Näen juoksevan miehen.	Case=Gen Number=Sing PartForm=Pres VerbForm=Part Voice=Act
Estonian	Näen juoksvat meest.	Case=Par Degree=Pos Number=Sing Tense=Pres VerbForm=Part Voice=Act
Hungarian	Látom a futó embert.	'ADJ'
Komi-Zyrian	Аддза котралысь мортӱс.	PartForm=Pres VerbForm=Part Voice=Act
(iv)	'Running is fun.'	
North Sámi	Viehkan lea suhtas.	Case=Nom Number=Sing
Erzya	Чїи́немаць вадря те́в.	Case=Nom Definite=Def Number=Sing VerbForm=Vnoun
Finnish	Juokseminen on kivaa.	Case=Nom Number=Sing
Estonian	Jooksmine on lahe.	Case=Nom Number=Sing
Hungarian	A futás jó dolog.	'NOUN'
Komi-Zyrian	Котравны лӱсьыд.	Case=Nom Number=Sing Tense=Past VerbForm=Part
(v)	'I like running.'	
North Sámi	Liikon viehkát .	VerbForm=Inf
Erzya	Вечка чїи́немам .	Case=Gen Number=Sing Number[psor]=Sing Person[psor]=1 Valency=2 VerbForm=Vnoun
Finnish	Pidän juoksemisesta .	Case=El Number=Sing
Estonian	Mulle meeldib joosta .	VerbForm=Inf
Hungarian	Szeretek futni .	VerbForm=Inf Voice=Act
Komi-Zyrian	Меным кажитчӱ котралӱм .	Case=Nom Number=Sing Tense=Past VerbForm=Part

Table 1: Table reproduced and adapted from Tyers and Pirinen (2016, p. 98)

The present participle in sentence (iii) is treated in all treebanks, except for Hungarian, as a deverbal form. Komi-Zyrian and Finnish deviate from these by introducing a 'PartForm=Pres' value, deviating from the 'Tense=Pres' strategy of the other treebanks.

In sentence (iv) the subject in nearly all languages is regularly a deverbal noun, although derivation or inflection is not indicated in the Hungarian treebank. While North Sámi, Finnish and Estonian use an attributive copula construction, Erzya and Hungarian apply an equation construction with a noun head, and Komi-Zyrian uses a simple infinitive to mark the primary argument with an attributive copula construction.

Sentence (v) provides multiple solutions for the second argument of the matrix verb. While the secondary argument in North Sámi, Hungarian and Estonian is an infinitive, the other languages use a deverbal noun. It should be noted that the deverbal noun is an object in Erzya, a subject in Komi-Zyrian and an oblique in Finnish. The subject–object dichotomy may be observed in the use of infinitives, as well.

5 Summary

At the moment, Uralic treebanks are mainly representative of the largest languages in each branch, although with the recent addition of a Karelian treebank the coverage is already spreading in the direction of smaller Finnic languages with a high representation of Balto-Finnic languages. In the same vein, smaller Sámi languages would be very welcome to UD, and similarly Udmurt and Moksha would increase the diversity of these branches. As mentioned previously, Mari and Samoyed treebanks are still missing

entirely, although some of these languages already have openly licensed annotated corpora that could easily be extended into treebanks. In principle the large coverage of Uralic languages at this point makes it realistic to expect that new treebanks would not introduce entirely different phenomena from what is already represented by the current Uralic languages. In the case of smaller and less studied Samoyedic languages, however, there may be questions that need specific attention, e.g. the expression of evidentiality and mood. Similarly Ob-Ugric languages may introduce dative alternation in Uralic languages.

The most common inconsistencies between languages in the Uralic treebanks seem to be related to the traditional terminology and concepts used in the description of individual languages. These are presumably the result of conversion schemes used when transforming different tagsets into UD. Especially with smaller treebanks improvements could be made relatively fast. However, since the inconsistency is large, it may not always be evident what the best shared solution is. The phenomena pointed out in this paper could be taken into account when systematizing the Uralic treebanks in future releases, although some of the work certainly falls beyond this language family into wider questions around cross-linguistic comparability in Universal Dependencies treebanks. One solution could be to create an explicit mapping between grammatically similar phenomena in the treebanks, and provide harmonization scripts that would adjust different phenomena into comparable representations. This could be connected to better documentation of the treebank conventions, ideally in a machine readable format, so that similar phenomena in different languages could be automatically linked to one another.

6 Acknowledgements

We want to thank three anonymous reviewers for their useful comments. Niko Partanen's work has been carried out within the project Language Documentation meets Language Technology: the Next Step in the Description of Komi, funded by the Kone Foundation.

References

- G. F. Fediunova. 2000. *Önija komi kyv, Morfologia*. Komi nebög ledzanin (Кomi небög лэдзанін). – Syktyvkar. ISBN 5-7555-0689-2.
- Ciprian Gerstenberger, Niko Partanen, Michael Rießler, and Joshua Wilbur. 2016. Utilizing language technology in the documentation of endangered Uralic languages. 4:29–47.
- Mika Härmäläinen and Jack Rueter. 2018. Advances in Synchronized XML-MediaWiki Dictionary Development in the Context of Endangered Uralic Languages. In *Proceedings of the Eighteenth EURALEX International Congress*, pages 967–978.
- Katri Haverinen, Jenna Nyblom, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Anna Missilä, Stina Ojala, Tapio Salakoski, and Filip Ginter. 2014. Building the essential resources for Finnish: the Turku Dependency Treebank. *Language Resources and Evaluation*, 48(3):493–531.
- Csilla Horváth, Norbert Szilágyi, Veronika Vincze, and Ágoston Nagy. 2017. Language technology resources and tools for Mansi: an overview. In *Proceedings of the Third Workshop on Computational Linguistics for Uralic Languages (IWCLUL 2017)*, pages 56–65.
- Mika Härmäläinen. 2019. UralicNLP: An NLP library for Uralic languages. *Journal of Open Source Software*, 4(37):1345.
- Boglarika Janurik. 2015. The emergence of gender agreement in code-switched verbal constructions in Erzya-Russian bilingual discourse. *Language Empires in Comparative Perspective*. De Gruyter, pages 199–218.
- Marja Leinonen. 2000. Evidentiality in Komi Zyryan. In Lars Johanson and Bo Utas, editors, *Evidentials: Turkic, Iranian and neighbouring languages*, pages 419–440. Walter de Gruyter.
- Sjur Moshagen, Jack Rueter, Tommi Pirinen, Trond Trosterud, and Francis M. Tyers. 2014. Open-Source Infrastructures for Collaborative Work on Under-Resourced Languages. In *The LREC 2014 Workshop “CCURL 2014 - Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era”*, pages 71–77.
- Kadri Muischnek, Kaili Müürisep, Tiina Puolakainen, Eleri Aedmaa, Riin Kirt, and Dage Särg. 2014. Estonian dependency treebank and its annotation scheme. In *Proceedings of 13th Workshop on Treebanks and Linguistic Theories (TLT13)*, pages 285–291.

Kadri Muischnek, Kaili Müürisep, and Tiina Puolakainen. 2016. Estonian Dependency Treebank: from Constraint Grammar tagset to Universal Dependencies. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Gabrielė Aleksandravičiūtė, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Arantza Diaz de Ilarraza, Carly Dickerson, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Tomaž Erjavec, Aline Etienne, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Johannes Heinecke, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Radu Ion, Elena Irimia, Olájidé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Hüner Kaşıkara, Andre Kaasen, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Arne Köhn, Kamil Kopacewicz, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phương Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Yuan Li, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Măranduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Tomohiko Morioka, Shinsuke Mori, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiack, Anna Nedoluzhko, Guntā Nešpore-Bērkalne, Luong Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Adédáyò Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Övrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Lapińska, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Daria Petrova, Slav Petrov, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Andriela Răăbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roșca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Abigail Walsh Sarah McGuinness, Dage Sārg, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamel Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Takaaki Tanaka, Isabelle Tellier, Guillaume Thomas, Liisi Torga, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Jing Xian Wang, Jonathan North Washington, Maximilian Wendt, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Daniel Zeman, Manying Zhang, and Hanzhi Zhu. 2019. Universal dependencies 2.4. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Niko Partanen, Rogier Blokland, KyungTae Lim, Thierry Poibeau, and Michael Rießler. 2018. The first Komi-Zyrian Universal Dependencies Treebanks. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 126–132.

Tommi Pirinen. 2019. Building minority dependency treebanks, dictionaries and computational grammars at

- the same time – an experiment in Karelian treebanking. In *Proceedings of the Third Workshop on Universal Dependencies (UDW 2019)*.
- Sampo Pyysalo, Jenna Kanerva, Anna Missilä, Veronika Laippala, and Filip Ginter. 2015. Universal dependencies for Finnish. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (Nodalida 2015)*, pages 163–172.
- Jack Rueter and Francis Tyers. 2018. Towards an open-source universal-dependency treebank for Erzya. In *Proceedings of the Fourth International Workshop on Computational Linguistics of Uralic Languages (IWCLUL 2018)*, pages 106–118.
- Mariya Sheyanova and Francis M. Tyers. 2017. Annotation schemes in North Sámi dependency parsing. In *Proceedings of the 3rd International Workshop for Computational Linguistics of Uralic Languages (IWCLUL 2017)*, pages 66–75.
- Francis M Tyers and Tommi A Pirinen. 2016. Intermediate representation in rule-based machine translation for the Uralic languages. In *Proceedings of the Second International Workshop for Computational Linguistics of Uralic Languages (IWCLUL 2016)*.
- Veronika Vincze, Katalin Simkó, Zsolt Szántó, and Richárd Farkas. 2017. Universal Dependencies and morphology for Hungarian - and on the price of universality. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 356–365, 01.
- Çağrı Çöltekin. 2016. (When) do we need inflectional groups? In *Proceedings of The First International Conference on Turkic Computational Linguistics*, pages 38–43.